Briefings in Bioinformatics, 2025, **26(2)**, bbaf085
https://doi.org/10.1093/bib/bbaf085
Case Study

Learning directed acyclic graphs for ligands and receptors based on spatially resolved transcriptomic data of ovarian cancer

Shrabanti Chowdhury[1,‡], Sammy Ferri-Borgogno[2,‡], Peng Yang[3], Wenyi Wang[3], Jie Peng[4], Samuel C Mok[2,*], Pei Wang[1,*]

[1]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, 1399 Park Ave, New York, NY 10029, United States
[2]Department of Gynecologic Oncology and Reproductive Medicine, Division of Surgery, The University of Texas MD Anderson Cancer Center, 1155 Pressler St., Houston, TX 77030, United States
[3]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, TX, United States
[4]Department of Statistics, University of California Davis, 399 Crocker Ln, Davis, CA 95616, United States

*Corresponding authors. Samuel C Mok, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, 1399 Park Ave, New York, NY 10029, United States. E-mail: scmok@mdanderson.org Pei Wang, Department of Gynecologic Oncology and Reproductive Medicine, Division of Surgery, The University of Texas MD Anderson Cancer Center, 1155 Pressler St., Houston, TX 77030, United States. E-mail: pei.wang@mssm.edu
‡Shrabanti Chowdhury and Sammy Ferri-Borgogno Co-first author.

Abstract

To unravel the mechanism of immune activation and suppression within tumors, a critical step is to identify transcriptional signals governing cell–cell communication between tumor and immune/stromal cells in the tumor microenvironment. Central to this communication are interactions between secreted ligands and cell-surface receptors, creating a highly connected signaling network among cells. Recent advancements in *in situ*-omics profiling, particularly spatial transcriptomic (ST) technology, provide unique opportunities to directly characterize ligand–receptor signaling networks that power cell–cell communication. In this paper, we propose a novel statistical method, `LRnetST`, to characterize the ligand–receptor interaction networks between adjacent tumor and immune/stroma cells based on ST data. `LRnetST` utilizes a directed acyclic graph model with a novel approach to handle the zero-inflated distributions of ST data. It also leverages existing ligand–receptor regulation databases as prior information, and employs a bootstrap aggregation strategy to achieve robust network estimation. Application of `LRnetST` to ST data of high-grade serous ovarian tumor samples revealed both common and distinct ligand–receptor regulations across different tumors. Some of these interactions were validated through both a MERFISH dataset and a CosMx SMI dataset of independent ovarian tumor samples. These results cast light on biological processes relating to the communication between tumor and immune/stromal cells in ovarian tumors. An open-source R package of `LRnetST` is available on GitHub at https://github.com/jie108/LRnetST.

Keywords: spatial transcriptomics data; ligand–receptor network; hill climbing; bootstrap aggregation; prior domain knowledge

Introduction

High grade serous ovarian cancer (HGSC) is the most lethal gynecologic malignancy, with its daunting overall survival rate showing limited improvement over decades [1–4]. A major obstacle in fully understanding the mechanisms of tumor progression and chemo-resistance in HGSC is its high intra-tumor heterogeneity, comprising both tumor clonal heterogeneity and tissue architecture heterogeneity [5]. The latter is reflected by the heterogeneous stromal and immune cell population in the ovarian tumor microenvironment (TME) [5, 6]. The recent advances in *in situ* omics analysis have suggested an important link between cell–cell interactions among tumor/immune/stromal cells in TME and tumor progression as well as therapeutic resistance [7]. However, the molecular mechanisms that shape these cell–cell interactions in HGSC are still largely unexplored.

A predominant form of cell–cell signaling is powered through interactions between ligands from one cell and cognate receptors on neighboring cells. Considerable effort has been dedicated to develop tools for exploring these interactions using single-cell RNA-seq (scRNA-seq) data, including CellphoneDB [8], SingleCellSignalR [9], Cellchat [10], ICELLNET [11], CrosstalkeR [12], Nichenet [13], scMLnet [14], and CytoTalk [15]. Despite these endeavors to characterizing intercellular communications with scRNAseq data, the absence of spatial information in scRNAseq data significantly hampers the precision in dissecting ligand–receptor (LR) interaction network, given that these interactions occur locally between neighboring cells within the tissue.

The latest development of spatial transcriptomic (ST) profiling technology enables mapping messenger RNA molecules to a specific location (spot or grid) of a tissue slice in a high-throughput manner [16–18]. These platforms thus provide an unprecedented opportunity to comprehensively characterize the LR interactions among neighboring cells (e.g. those from the adjacent grids on ST slices), which is not feasible based on either bulk or single-cell RNA profiles. In this paper, we aim to characterize the LR regulatory network in HGSC using ST data.

Received: October 8, 2024. **Revised:** January 29, 2025. **Accepted:** February 17, 2025

© The Author(s) 2025. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Pioneering efforts have been made for inferring cell–cell communication networks while considering spatial information, including Giotto [19], stLearn [20], and spaCI [21]. Within Giotto, the Spatially Informed Cell-to-Cell Communication (*spatCellCellcom*) module calculates cell–cell communication scores for each LR pair between proximal cell types according to the spatial network. Permutation-based *P*-values and multiple hypotheses adjustment are subsequently computed. stLearn focuses on the proportion of neighboring spots with upregulated expressions for both the ligand and receptor genes of a given LR pair. The significant LR pairs are then obtained by integrating the signals across all spots. spaCI uses the gene-based and spatially guided embeddings to convert the gene expressions into latent representation via a standard multilayer perceptron and then applies a triplet loss function to predict the cosine similarity of all possible LR pairs. Although all three methods incorporate spatial information, none directly addresses the issue of zero-inflation that is pervasive in ST data. Furthermore, assessing co-expression based on either thresholding, marginal correlation, or similarity measure is susceptible to considerable variability present in ST data, leading to a lack of reproducibility (see Results section).

To address these challenges, we developed LRnetST—Ligand-Receptor Network learning based on Spatial Transcriptomics data, a novel tool to construct LR networks between adjacent cells of different types based on either multicellular or single-cell ST data. The LRnetST pipeline begins by introducing the Neighbor Integrated Matrix (NIM), which integrates the spatial information and the molecular information within the ST data. Subsequently, LRnetST utilizes a binary variable alongside a continuous variable for every ligand/receptor in the node space. This coding strategy not only addresses zero inflation in ST data, but also enhances the power to detect interactions predominantly signaled through active/inactive statuses of ligands/receptors. To account for variation in library sizes across grids/cells, LRnetST employs an aggregation framework that combines bootstrap resample based directed acyclic graph (DAG) learning with downsampling based normalization. This framework provides false edge control and is adaptable to incorporate prior information, e.g. data from existing LR databases.

We applied LRnetST to both multicellular and single-cellular ST datasets of ovarian cancer. In the multicellular dataset, our focus was on detecting LR interactions between neighboring spot-pairs enriched with tumor and immune/stroma cells, respectively. In the single-cellular ST data, we examined interactions between tumor cells and other immune or stroma cell types. To guide the construction of DAGs, we incorporate known LR regulation information from relevant databases [22] as prior information to constrain the DAG space. Based on the multicellular analysis, LRnetST revealed a substantial number of shared LR interactions between tumor and stroma cells across four different HGSC samples. Further, applying LRnetST to the single-cell MERFISH ST data as well as CosMx SMI ST data of independent ovarian tumors, we were able to confirm several LR interactions between tumor and nearby stroma cells. These findings offer fresh insights into the roles of these tumor-relevant genes/proteins in facilitating cell–cell communication within HGSC.

Materials and methods

Notation and background

A DAG $\mathcal{G}(\mathbb{N}, \mathbb{E})$ contains a node set $\mathbb{N}$ and an edge set $\mathbb{E}$, where the edges are directed from *parent* nodes to *children* nodes, without any cycle in the graph. In a DAG model, the node set $\mathbb{N}$ represents a set of random variables, while the edge set $\mathbb{E}$ represents the conditional dependence relationships among these random variables. Structure learning of a DAG means identifying the parent set (often known as neighborhood) of each node in the graph.

LRnetST pipeline

To effectively characterize the complicated LR regulatory relationship in TME, we proposed to use DAG models, which has emerged as a valuable tool for inferring gene–gene interactions [22–28]. However, multiple challenges impede the application of existing DAG models to ST data, including their inability to integrate both spatial information and -omics profiles concurrently into network learning, statistical and computational challenges due to high drop-out rates in the ST data, and read count variation across different ST spots (see Methods).

To address these challenges, we developed a new DAG model coupled with customized fitting pipeline, called LRnetST, for constructing LR networks based on ST data. Below, we first provide a concise overview of the pipeline, followed by a detailed explanation of the key steps.

As illustrated in Fig. 1A, the LRnetST pipeline comprises of four major steps: (1) normalization through downsampling, (2) construction of the NIMs with an additional round of bootstrap resample, (3) estimation of DAGs based on bootstrapped samples, and (4) aggregation of the DAGs into a final network.

Within the LRnetST pipeline, the second step helps to combine the spatial location information with the molecular data information. It contains three key components as illustrated in Fig. 1B: (1) integration of the spatial and molecular information through introducing the Initial Neighbor Integrated Matrices (Initial-NIMs), (2) forming the NIM by encoding each gene expression using one binary indicator and a continuous variable, and (3) bootstrap resampling for the subsequent DAG construction. Notably, this framework, which integrates downsampling-based normalization, NIM construction, and the bootstrap-aggregation based network inference, not only effectively accounts for the varying unique molecular identifier (UMI) count levels across different ST spots/cells, but also enhances the robustness of the DAG structure learning.

The third step of LRnetST contains a novel DAG model customized to NIM. Specifically, the DAGs are constructed by maximizing a Bayesian information criterion (BIC) penalized joint log-likelihood of all the nodes in the graph. Moreover, LRnetST incorporates existing LR databases as prior information to constrain the search space during DAG learning.

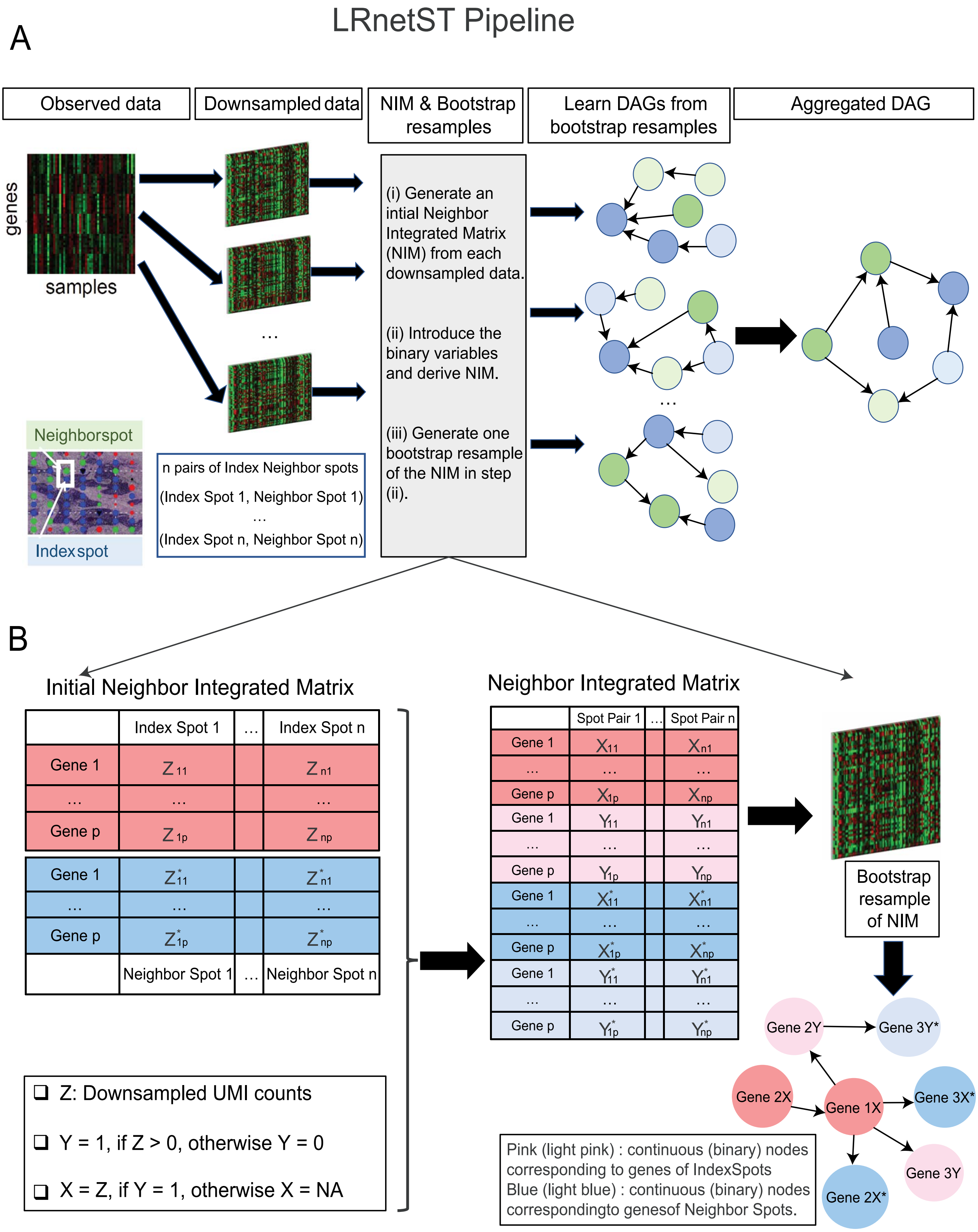
In the last step, the derivation of the final aggregated DAG is achieved by searching for a DAG that minimizes an average structural Hamming distance (SHD) to the DAGs in the ensemble built from the previous steps.

Note that LRnetST is a versatile tool and can be applied to both multicellular and single-cellular ST datasets. In the real data application, we demonstrated LRnetST's utility by analyzing both a multicellular and a single-cell ST datasets of ovarian cancer. Primarily the multicellular data were used to identify LR interactions between neighboring spot-pairs enriched with tumor and stroma cells, respectively, while two independent single-cell ST datasets were used to validate the findings.

Below, we provide details of the key steps in the LRnetST pipeline using multicellular ST data. The parallel strategy for handling single-cell ST data is similar, with "spots" replaced by "cells."

Neighbor integrated matrix

In the LRnetST pipeline, we first derive an Initial-NIM to integrate the spatial and the molecular information in ST data. We choose



n pairs of Index Neighbor spots
(Index Spot 1, Neighbor Spot 1)
…
(Index Spot n, Neighbor Spot n)

(i) Generate an intial Neighbor Integrated Matrix (NIM) from each downsampled data.

(ii) Introduce the binary variables and derive NIM.

(iii) Generate one bootstrap resample of the NIM in step (ii).

	Index Spot 1	…	Index Spot n
Gene 1	Z_{11}		Z_{n1}
…	…		…
Gene p	Z_{1p}		Z_{np}
Gene 1	Z^*_{11}		Z^*_{n1}
…	…		…
Gene p	Z^*_{1p}		Z^*_{np}
	Neighbor Spot 1	…	Neighbor Spot n

❑ Z: Downsampled UMI counts

❑ Y = 1, if Z > 0, otherwise Y = 0

❑ X = Z, if Y = 1, otherwise X = NA

	Spot Pair 1	…	Spot Pair n
Gene 1	X_{11}		X_{n1}
…	…		…
Gene p	X_{1p}		X_{np}
Gene 1	Y_{11}		Y_{n1}
…	…		…
Gene p	Y_{1p}		Y_{np}
Gene 1	X^*_{11}		X^*_{n1}
…	…		…
Gene p	X^*_{1p}		X^*_{np}
Gene 1	Y^*_{11}		Y^*_{n1}
…	…		…
Gene p	Y^*_{1p}		Y^*_{np}

Pink (light pink) : continuous (binary) nodes corresponding to genes of IndexSpots
Blue (light blue) : continuous (binary) nodes correspondingto genesof Neighbor Spots.

Figure 1. LRnetST pipeline. (A) Given the ST data matrix and the spatial information, LRnetST generates multiple downsampled data matrices together with a list of matching Index Spot and Neighbor Spot pairs, constructs NIMs with bootstrap resamples, learns DAGs from resampled NIMs, and aggregates the inferred DAGs using SHD. (B) To construct one DAG, starting with the initial NIM by stacking the gene expression profiles of the matching Index and Neighbor Spots, LRnetST converts the downsampled UMI counts Z into activation status Y and expression level X given Y, generates a bootstrap resample without replacement, and learns a DAG based on the resampled matrix.

to focus on pairs of adjacent ST spots enriched of tumor and stromal/immune cells, respectively. Specifically, for each tumor sample, ST spots are first classified into two classes: enriched or not enriched of tumor cells. We then identify the subset of tumor-cell-enriched spots sitting on the boundary of the tumor regions (i.e. being adjacent to spots enriched of non-tumor cells on the ST slice). We refer to these spots as the Index Spots and refer to their closest non-tumor-cell-enriched spots as their Neighbor Spots. We denote the total number of Index Spot and Neighbor Spot pairs as n, and the number of genes under consideration as p.

In the next step, we derive an Initial-NIM by stacking the rows of the gene expression matrix of the Index Spots, which has a dimension of $p \times n$, with the gene expression matrix of the corresponding Neighbor Spots, which also has a dimension of $p \times n$. The resulting Initial-NIM has an expanded feature space of $2p$ rows. Finally, we convert the *Initial-NIM* into *NIM* by replacing the expression Z of a gene with a pair of features, a continuous variable X and a binary variable Y, to facilitate the modeling of zero-inflation in the ST data (see the next subsection). The resulting NIM has a dimension $4p \times n$. See Fig. 1B and the pseudo algorithm outlined in section A.2 of the Supplementary Material for more details.

Accounting for zero inflation in ST data

Similar to scRNA-seq data, ST data have a high dropout rate, i.e. only a fraction of the transcriptome detected at each ST spot. To facilitate the modeling of such zero-inflated data, `LRnetST` uses two nodes (variables) to represent each gene: a binary node and a continuous node.

First, we use Z_{ij} to denote the normalized expression measures (see the next subsection) of the i_{th} gene in the j_{th} ST spot. Note that $Z_{ij} = 0$ corresponds to a situation where either the i_{th} gene is not expressed, or its expression level is low in the j_{th} spot and thus is not detected in the sequencing experiments. Therefore, we view the event of $Z_{ij} = 0$ as a "less-active" status of the gene and uses a binary node, denoted as Y_{ij}, to represent the "detection status":

$$Y_{ij} = 1, \text{ if } Z_{ij} > 0 \text{ (i.e. the gene is detected); and}$$

$$Y_{ij} = 0, \text{ otherwise.}$$

Given the gene is detected. (i.e. $Y_{ij} = 1$), the continuous node, denoted as X_{ij}, represents the normalized gene expression/abundance:

$$X_{ij} = Z_{ij}, \text{ if } Y_{ij} = 1; \text{ and } X_{ij} = NA, \text{ if } Y_{ij} = 0.$$

By including both nodes, X_{ij} and Y_{ij}, `LRnetST` not only accounts for the zero inflation in the data, but also achieves better power in detecting those interactions that are largely signaled through the active/less-active statuses of genes.

Score function and optimization

Based on NIM, we learn an LR DAG by minimizing a BIC-type score function through a greedy search algorithm—*hill-climbing* (HC) [25, 26, 28, 29]. Note that the likelihood of the continuous nodes is calculated only using data from spots on which their binary nodes take the value 1 (i.e. X is not NA). This makes the binary nodes natural parents of the corresponding continuous nodes of the same gene. At each current graph $\mathcal{G}$, for a continuous node, X, the residual sum of squares ($RSS_X^{\mathcal{G}}$) from regressing X onto its current parent set, using only the samples where X is nonzero, is calculated. Specifically, we have $loglik(X|pa_X^{\mathcal{G}}) \sim -n_X/2\,\mathbf{log}(RSS_X^{\mathcal{G}}/n_X) + Constant$, where, $RSS_X^{\mathcal{G}}$ is the residual sum of squares, $pa_X^{\mathcal{G}}$ denotes the parent set of the node in graph $\mathcal{G}$, and n_X is the number of samples where X is nonzero. The log-likelihood of a binary node Y is obtained by regressing Y onto its current parent set $pa_Y^{\mathcal{G}}$ through logistic regression [29].

$$loglik(Y|pa_Y^{\mathcal{G}}) = \left(\sum_{s=1}^{n} I(Y_s = 1)\,\mathbf{log}(\hat{p}_s) + \sum_{s=1}^{n} I(Y_s = 0)\,\mathbf{log}(1 - \hat{p}_s)\right),$$

where $\hat{p}_s = \widehat{P}\left(Y_s = 1|pa_{Y,s}^{\mathcal{G}}\right) = \frac{\mathbf{exp}(\hat{\alpha}_0 + \hat{\boldsymbol{\alpha}}^T pa_{Y,s}^{\mathcal{G}})}{1+\mathbf{exp}(\hat{\alpha}_0 + \hat{\boldsymbol{\alpha}}^T pa_{Y,s}^{\mathcal{G}})}$, $\hat{\alpha}_0$ is the fitted intercept and $\hat{\boldsymbol{\alpha}}$ is the vector of the fitted coefficients, and n is the sample size. We then consider BIC scores that penalize for model complexity: $score_{BIC_X} = -2loglik(X|pa_X^{\mathcal{G}}) + |pa_X^{\mathcal{G}}|\,\mathbf{log}(n_X)$ for a continuous node X and $score_{BIC_Y} = -2loglik(Y|pa_Y^{\mathcal{G}}) + |pa_Y^{\mathcal{G}}|\,\mathbf{log}(n)$ for a binary node Y, where $|pa_X^{\mathcal{G}}|$ and $|pa_Y^{\mathcal{G}}|$ denote the size of the parent set of a continuous and a binary node, respectively, in graph $\mathcal{G}$.

The final score of a graph $\mathcal{G}$ is the summation of the BIC scores of the individual nodes. We then search for the DAG that minimizes this score function using an efficient implementation of the HC algorithm that is modified from our previous work `DAGBagM` [29].

Downsampling normalization and bootstrap model aggregation

`LRnetST` employs an aggregating framework that couples a bootstrap aggregation (bagging) procedure of DAG learning with the downsampling based normalization to achieve better control of false edge detection and at the same time to account for the varying library sizes across different ST spots.

Specifically, we apply downsampling normalization on the original gene expression data by fixing the total UMI at the median level across all ST spots. We then generate $B = 100$ downsampled gene expression matrices ([Z]), and construct $B = 100$ NIMs (see Fig. 1B) accordingly. Next we obtain one bootstrap resample on each NIM through sampling with replacement of the Index/Neighbor Spots pairs. Finally, we learn one DAG based on each bootstrap resample using the method described in the previous subsection.

The resulting ensemble of DAGs are then aggregated following the aggregation procedure implemented in `DAGBagM` [29], where the HC algorithm is (again) used to search for a DAG that minimizes an aggregation score based on the *SHD* while maintaining acyclicity.

For more details of the `LRnetST` pipeline, please see the pseudo code in Algorithm in Section A.2 of the Supplementary Material.

Simulation studies

We utilized synthetic data sets to evaluate the performance of the DAG learning step in `LRnetST`. We consider simulations with different sample sizes (i.e. numbers of Index-Neighbor Spot pairs: $n = 200$ and 400) and different DAG topology with varying number of genes ($p = 100$ to 600). The true DAGs are shown in Figure S1A-F. For the detailed data generating steps please refer to Section A.3 of the Supplementary Material. We compared the performance of DAG learning based on initial-NIM and based on NIM to assess the advantage of using the paired binary and continuous nodes to facilitate the modeling of zero-inflation in ST data. Moreover, we compared the performance of `LRnetST` with three alternative methods: `DAGBagM`, `DAGBagM_C` [29], and `bnlearn` [30]. `DAGBagM` is a method for learning DAGs with mixed binary and continuous nodes. When applying `DAGBagM` to NIM, X is replaced with Z (i.e. NA in X is replaced with 0) in NIM, and all samples are used to calculate the scores for the continuous nodes, whereas LRnetST uses the converted X so that only samples with $Y = 1$ are used to calculate the scores for the continuous nodes. Note that we performed bootstrap aggregation for all methods except bnlearn due to its prohibitive computational cost. To evaluate the performance of LRnetST and other methods for detecting the skeleton (i.e. edges without direction) as well as the directed edges, we assessed the power (or true positive rates), false discovery rates, and the F1 scores defined as $F1 = 2 \times \text{precision} \times \text{recall}/(\text{precision}+\text{recall})$.

ST data description and processing

10x ST data: to characterize tumor and its microenvironment, ST analysis was performed on four fresh frozen treatment-naive advanced stage HGSC samples using the 10x ST platform as described in our previous work [31]. Among these samples, two (A4 and A5) were derived from chemo-sensitive patients with extended progression free survival time (> 10 years), and the remaining two (A10 and A12) originated from chemo-resistant patients with short progression free survival time (< 6 months). In the ST experiment, each ST spot contained 20–50 cells, and 930–1007 spots were profiled per tumor slice.

Expression levels of genes in each spot were measured using counts of UMI. Since each spot is a mixture of several cell types, in order to identify tumor or immune/stromal cell-enriched spots, we estimated the cell-type proportions in each grid (spot) by performing deconvolution analysis. We then integrated the estimated cell-type proportions with pathologist-annotated tumor/stroma regions in each tumor slice to identify the tumor-enriched spots, as well as spots enriched with immune/stromal cells (see Section A.4 in the Supplementary Material). For each tumor spot, we further identified the adjacent or neighboring immune/stromal spots that are at most 2-grid distance apart from the tumor spot.

MERFISH and CosMx SMI ST data: to validate the LR edges from 10x ST data, we obtained publicly available MERFISH data sets for four ovarian tumor samples [32] and a NanoString CosMx Spatial Molecular Imaging (SMI) dataset of one HGSC tumor [33].

After preprocessing, for each tumor, we performed clustering and cell-type annotation analysis (see Supplementary Materials). For MERFISH, we identified five major cell groups: tumor cells, macrophages, T-cells, fibroblast, and endothelial, while for CosMx SMI, we identified tumor, macrophages, and T-cells as the major cell groups (Supplementary Table S2). Then, for each tumor cell, we identified its adjacent/neighboring cell that has minimum distance from the tumor cell, as obtained from the Delaunay network (implemented under Giotto).

Normalization: for all the 10x, MERFISH, and CosMx SMI data, the median of the total gene counts per spot (or cell) in each sample was used for the downsampling normalization in LRnetST. Tables A.3 and A.4 in the Supplementary Material give the detailed numbers of the tumor and neighboring cells, the numbers of adjacent index-neighbor-cell pairs, along with the number of LR genes (documented in database [34]) used for LR network learning in each tumor tissue for 10x ST and MERFISH data, respectively.

Comparing LRnetST with stLearn, spatCellCellcom, and spaCI

To compare the performance of LRnetST with other LR network learning methods, we applied spatCellCellcom, stLearn, and spaCI on the same data sets. For preprocessing including normalization, we followed their respective pipelines implemented in the corresponding R (spatCellCellcom implemented under Giotto) and python (stLearn and spaCI) packages. Similar to LRnetST, we only considered the genes documented in the LR database. For spatCellCellcom and stLearn, we obtained the LR regulation edges (Table A.5 in the Supplementary material, Supplementary Table S1) by thresholding the adjusted P-values, while for spaCI we used a threshold on the predicted cosine similarity from the output of all possible LR interactions to derive networks of comparative number of edges.

Results

Evaluation of DAG learning based on synthetic data sets

We evaluated and compared the performance of the DAG learning step in LRnetST with alternatives including DAGBagM [29], DAGBagM_C [29], and bnlearn, based on a collection of simulated NIM matrices (see details in Methods). The results in Table A.1 in the Supplementary Material and Figure S1G-H indicate superior performance of LRnetST compared with the other methods across all considered scenarios, demonstrating the results of including the binary nodes in NIM in the LRnetST pipeline. Furthermore, the results of LRnetST and DAGBagM on NIM show the advantages derived from the customized treatment of likelihood terms associated with paired binary and continuous nodes in LRnetST. In the end, as expected, for all four methods, we observed improved performances as the sample size (n) increases, while declined performance with the increase of the number of genes (p).

Revealing cell–cell interaction in ovarian cancer with LRnetST

Application to the 10× ST ovarian cancer data

For each tumor sample in the ovarian cancer 10x ST data set, we applied LRnetST, stLearn spatCellCellcom, and spaCI to construct patient-specific LR networks (Figure S2A–D). Table A.5 in the Supplementary Material summarizes the numbers of LR regulation edges.

Reproducibility of the edges across four tumor samples

We first evaluated how ”reproducible” the inferred LR edges are across different tumor samples. Intuitively, methods detecting more shared (reproducible) edges across tumors shall enjoy better power for detecting meaningful biological interactions. As summarized in Fig. 2A,C–E, LRnetST inferred many more shared (reproducible) edges across multiple (at least two) samples than the other three methods. Specifically, LRnetST inferred 185 reproducible edges, while spatCellCellcom inferred 67, stLearn inferred 48, and spaCI inferred only 21 reproducible edges. Note that, for counting the reproducible edges, we considered undirected edges and further collapsed the binary and continuous components of each gene to one node. Further, assessing whether the overlap between LR edges for a pair of tumors can significantly surpass that by random chance (Fig. 2B), we found the results of LRnetST are significant for all six pairs of tumors, whereas the results from spatCellCellcom, stLearn, and spaCI demonstrated significant overlap for no more than half of the pairs. Finally, LRnetST identified four edges shared by all four tumors, whereas the other methods failed to identify any such edges. These results suggest a higher level of reproducibility of LRnetST compared with the alternatives.

We then investigated the hub nodes in the LR networks by the four methods (Figs S3, S4). Hub nodes are those with higher connections in a network, and thus are often more likely to play an essential role [35]. We identified three common hub nodes, LRP1, ITGB1, and ITGAV, across all four tumors by LRnetST. On

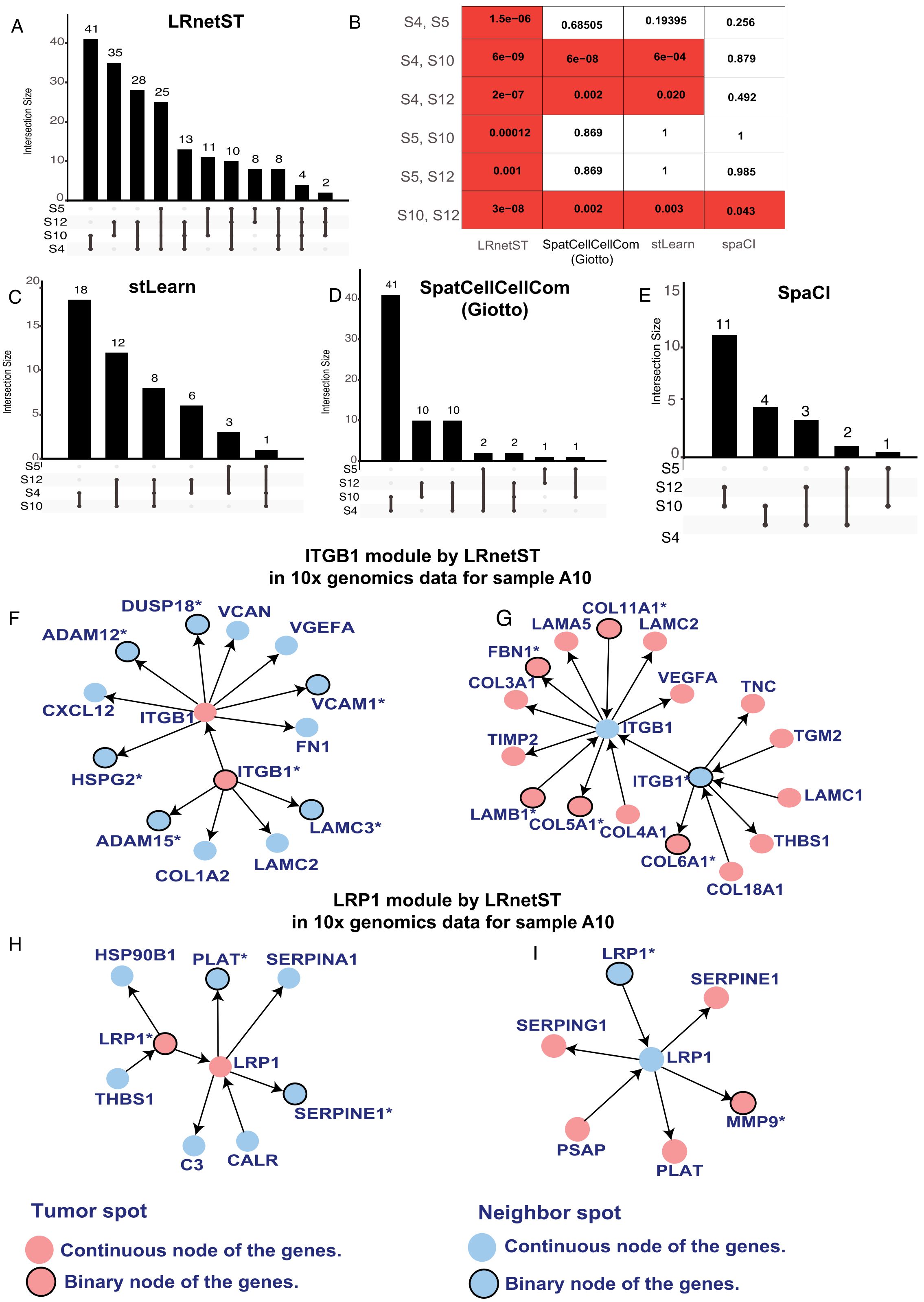


Figure 2. Reproducible LR edges and LR network modules resulted from the 10x ST data. (A) Common edges inferred by LRnetST across the 4 10x ST data. (B) *P*-values for the pairwise correlation of the common edges between the samples by LRnetST, spatCellCellcom (Giotto), stLearn, and spaCI. (C) Common edges inferred by stLearn across the 4 10x ST data. (D) Common edges inferred by spatCellCellcom (Giotto) across the 4 10x ST data. (E) Common edges inferred by spaCI across the 4 10x ST data. (F)One ITGB1 module by LRnetST in 10x ST data sample 10, where ITGB1 in the tumor spot is causally associated with several genes from the neighbor spot. (G) The otherITGB1 module by LRnetST in 10x ST data sample 10, where ITGB1 in the neighboring stroma spot is causally associated with several genes from the tumor spot. (H) One LRP1 module by LRnetST in 10x ST data sample 10, where LRP1 in the tumor spot is causally associated with several genes from the neighbor spot. (I) The other LRP1 module by LRnetST in 10x ST data sample 10, where LRP1 in the neighboring stroma spot is causally associated with several genes from the tumor spot.

the other hand, only ITGB1 was detected as a common hub node by spatCellCellcom, stLearn, and spaCI.

Interestingly, by LRnetST (Fig. S3A), LRP1 in both tumor cell-enriched spots (LRP1_tumor) and stroma cell-enriched spots (LRP1_stroma) appeared to be a hub node in all LR networks. Of all the edges associated with LRP1, the interaction between LRP1_stroma and MMP9_tumor is the only common edge identified in all four tumor samples (Fig. 2H and Fig. S7B, D, F). Previous studies have highlighted the active involvement of the LRP1–MMP9 interaction in many biological processes [36] (see Discussion).

Moreover, a few interactions involving LRP1 were patient-specific, such as LRP1_Tumor–C1QB_Stroma interaction was detected only in the two chemo-sensitive patients, whereas, LRP1_Tumor–SERPINA1_Stroma was detected only in the two chemo-refractory patients (Fig. 2G-H, and Fig. S7A–F). The latter implies a potential connection between LRP1 and the protease [37].

Apart from LRP1, ITGB1 is another hub node inferred to interact with multiple genes in all four tumors by LRnetST (Fig. 2F and G, and Fig. S8A–D). Specifically, for both chemo-refractory patients, ITGB1_tumor was detected to interact with VCAM1 and VEGFA from the nearby stroma spots. At the same time, ITGB1_stroma was connected with multiple genes, including LAMC2 and VEGFA, from the nearby tumor spots. These interactions have been previously reported in multiple literatures [38, 39] (see Discussion).

To further assess the performance of various methods, we then applied LRnetST and the other three methods to two independent cell-level ST data of ovarian cancer, one from the MERFISH platform and the other from the NanoString CosMx SMI platform, to validate some of the detected LR interactions.

Application to MERFISH and CosMx SMI data for validation

We studied four MERFISH data sets [32] derived from four tumor slices of two ovarian cancer patients (Methods). The MERFISH datasets contain gene expression measurements of 550 genes in over 100 000 single cells. The CosMx SMI data contain gene expression measurements for a panel of 960 genes in 21 651 single cells from a HGSC patient.

Both the MERFISH and the CosMx SMI data offer cell-level resolution, enabling detailed monitoring of LR interactions between individual cells. This granularity is particularly advantageous for overcoming the challenge of cell type mixtures within spots encountered in 10x ST data. However, owing to technology limitations inherent to MERFISH and CosMx SMI, only a limited set of transcripts (few hundreds) can be detected and quantified. This significantly constrains the ability to systematically characterize LR interactions, underscoring our use of MERFISH and CosMx SMI data for validation rather than exploration purposes.

For preprocessing and clustering of the validation datasets we used the same pipeline implemented under R Giotto (see section A.4 of Supplementary material).

Figures 3A and 4A illustrate the spatial distribution of tumor cells and macrophage cells in one tumor slice (see Fig. S9A–C for other cell-types) for MERFISH and CosMx SMI data, respectively. While there are thousands of epithelial (pink) and macrophage or T-cells (light blue) cells in the tumor slice, we focused on the neighboring tumor-macrophage (or tumor-T-cells) pairs (red epithelial cells and dark blue macrophage cells). With LRnetST, we built a collection of LR networks, one for each cell type pair and each tumor slice for both validations datasets (Supplementary Tables S3 and S4). In the result of the MERFISH data, there were 53 nodes involved for LR interactions that were inferred in at least two tumor slices. Figure 3B–E illustrates subsets of inferred networks considering different neighboring cell type pairs for one tumor slice. Similarly, a subset of the inferred LR networks from the CosMx SMI data are shown in Fig. 4B and C, considering neighboring tumor-macrophage and tumor-T-cells pairs.

We then assessed the validation of the inferred LR interactions from the 10x ST data. Specifically, we focused on the 185 reproducible edges identified in at least two out of the four 10x ST data sets by LRnetST. Among these, only 70 edges had both nodes measured in the MERFISH and/or CosMx SMI datasets. Of these, 30 edges were detected in the LR network constructed from either the MERFISH or CosMx SMI data (Fig. 5A, Table A.6 in the Supplementary Material). Additionally, 10 edges were confirmed based on both datasets (bold red and bold green edges in Fig. 5A). For example, the interaction between ITGB1_tumor and VCAM1_stroma as well as VEGFA_stroma were inferred in multiple tumor samples from all three data sets (Supplementary Tables S3 and S4), suggesting the robust signal of these LR interactions in ovarian tumor tissues.

In parallel, we performed the same analysis using spatCellCellcom, stLearn, and spaCI (Fig. 5B and C, Supplementary Tables S3 and S4). As illustrated in Fig. 5B and C, only three and one edges were validated in the MERFISH data, while two and three other edges were validated in CosMx SMI data, by stLearn and SpatCellCellcom, respectively. Note that no edge was inferred in all three data sets by either spatCellCellcom and stLearn (Fig. 5B and C). For spaCI, we obtained sparse LR networks and only the interaction between ITGB1 and VEGFA was validated (Supplementary Table S4; no Figure provided)). Note that one interaction between ITGB1 and VEGFA was detected and validated by all four methods.

Discussions

To gain insights on the molecular basis of TME, in this paper, we introduce a new method—LRnetST—to infer LR interaction networks among adjacent cells using ST data. LRnetST employs novel strategies to address the spatial structure and high dropout rates that are uniquely present in the ST data. We demonstrated benefits of LRnetST for modeling the zero-inflated ST data over alternative approaches on synthetic data.

Furthermore, we compared performance of LRnetST and alternative tools on the ovarian cancer ST datasets. The results highlight LRnetST's superior capability in detecting reproducible signals that hold greater biological significance. However, similar to other high-dimensional network construction methods, the performance of LRnetST may be impacted by the dimension of the gene space, as demonstrated in our simulation study. This underscores the significance of leveraging the existing LR interaction databases as prior domain knowledge to constrain the network space, as exemplified in our ovarian cancer ST data analysis.

By LRnetST, LRP1 is identified as a hub node in LR networks. LRP1 is reported to make critical contribution to many processes that drive tumorigenesis and tumor progression [40, 41], and has been proposed to be an important diagnostic and prognostic biomarker of epithelial ovarian cancer [36]. Besides the previously reported role of LRP1 in ERK signaling pathway through interaction with MMP9 [36], which was also supported in our results (Fig. 2G), we further revealed strong interactions between LRP1 in tumor cells and SERPINA1 (AAT) (Fig. 3B and C, 5A, and S7A, S7C, S7E) in the two short-term survivor samples in the 10x ST data

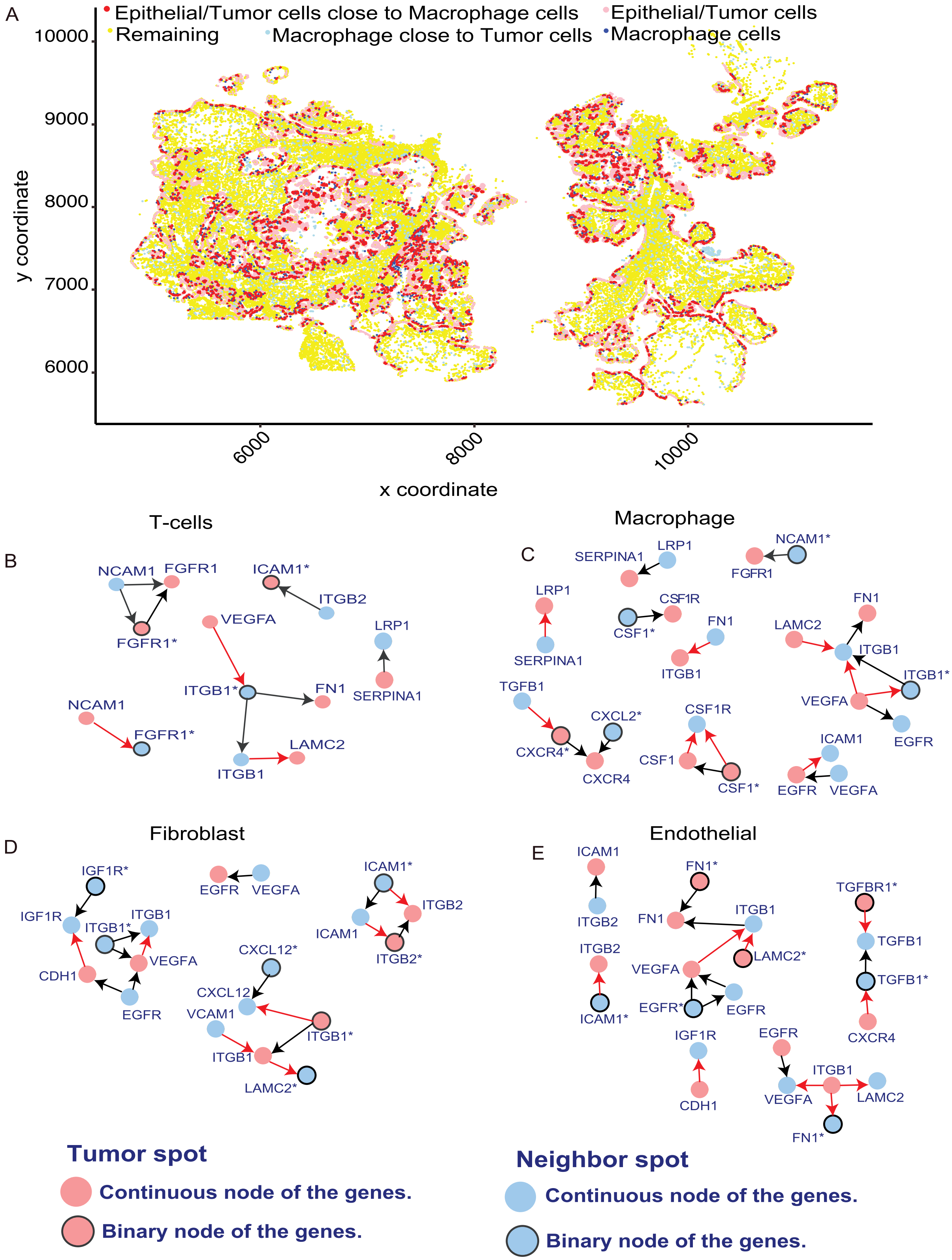


Figure 3. Inferred LR network based on the MERFISH data (Patient-2, slice-2) by LRnetST. (A) Spatial plots showing the index epithelial (tumor) cells in red, the neighboring macrophage cells close to the tumor cells in blue, the non-index epithelial (tumor) cells in pink, and the non-adjacent macrophage cells in light blue. (B)–(E) LR network topology for LR interactions inferred in at least two MERFISH data sets, with red edges indicating those detected also in the 10x ST data.

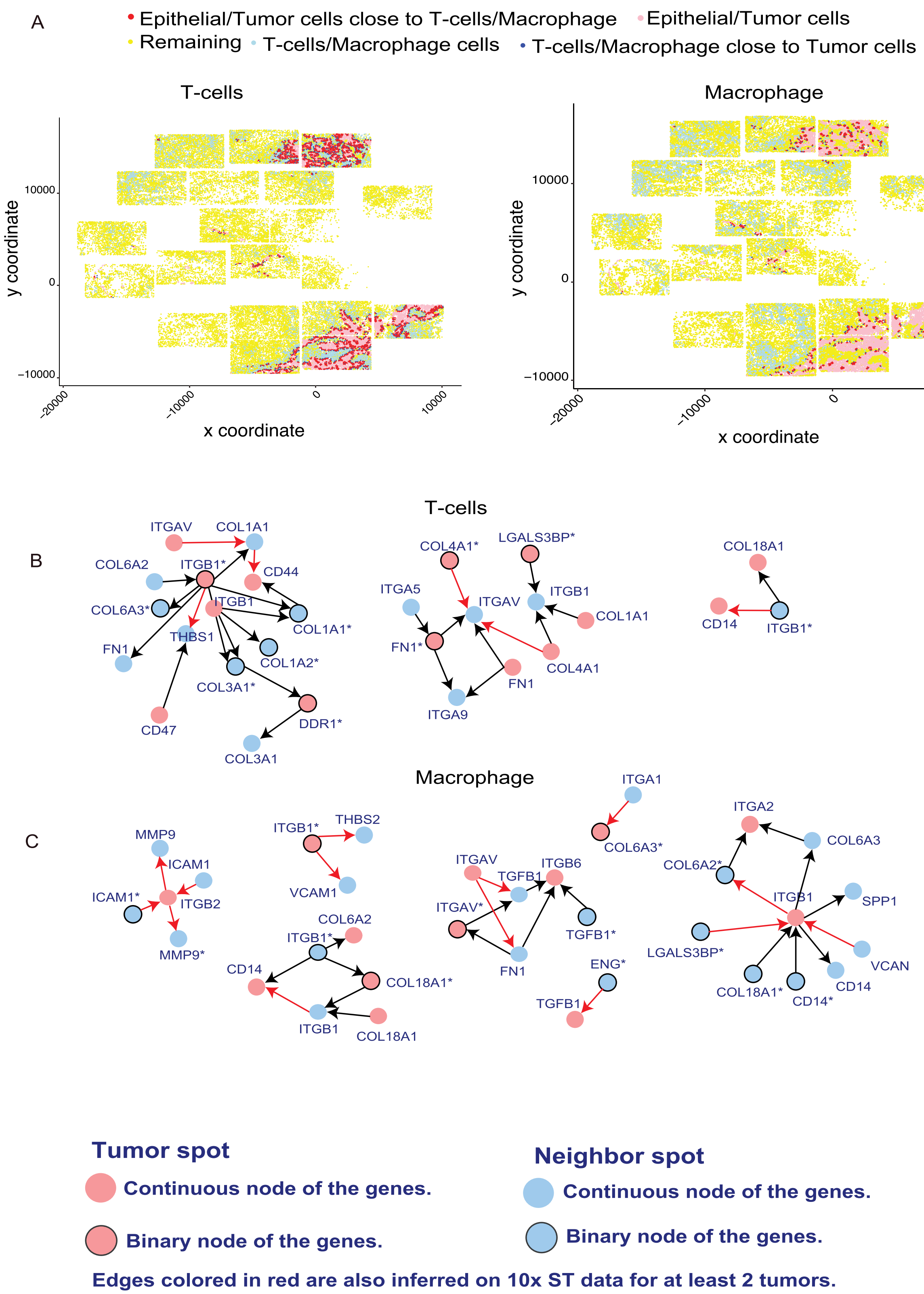


Figure 4. Inferred LR network based on Nanostring CosMx SMI data by LRnetST. (A) Spatial plots showing the index epithelial (tumor) cells in red, the neighboring T-cells (left) or macrophage (right) cells close to the tumor cells in blue, the non-index epithelial (tumor) cells in pink, and the non-adjacent macrophage cells in light blue. (B) and (C) LR network topology for modules with at least one LR interaction appearing in two or more LR networks from the 10x ST data, with red edges indicating those detected also in the 10x ST data.

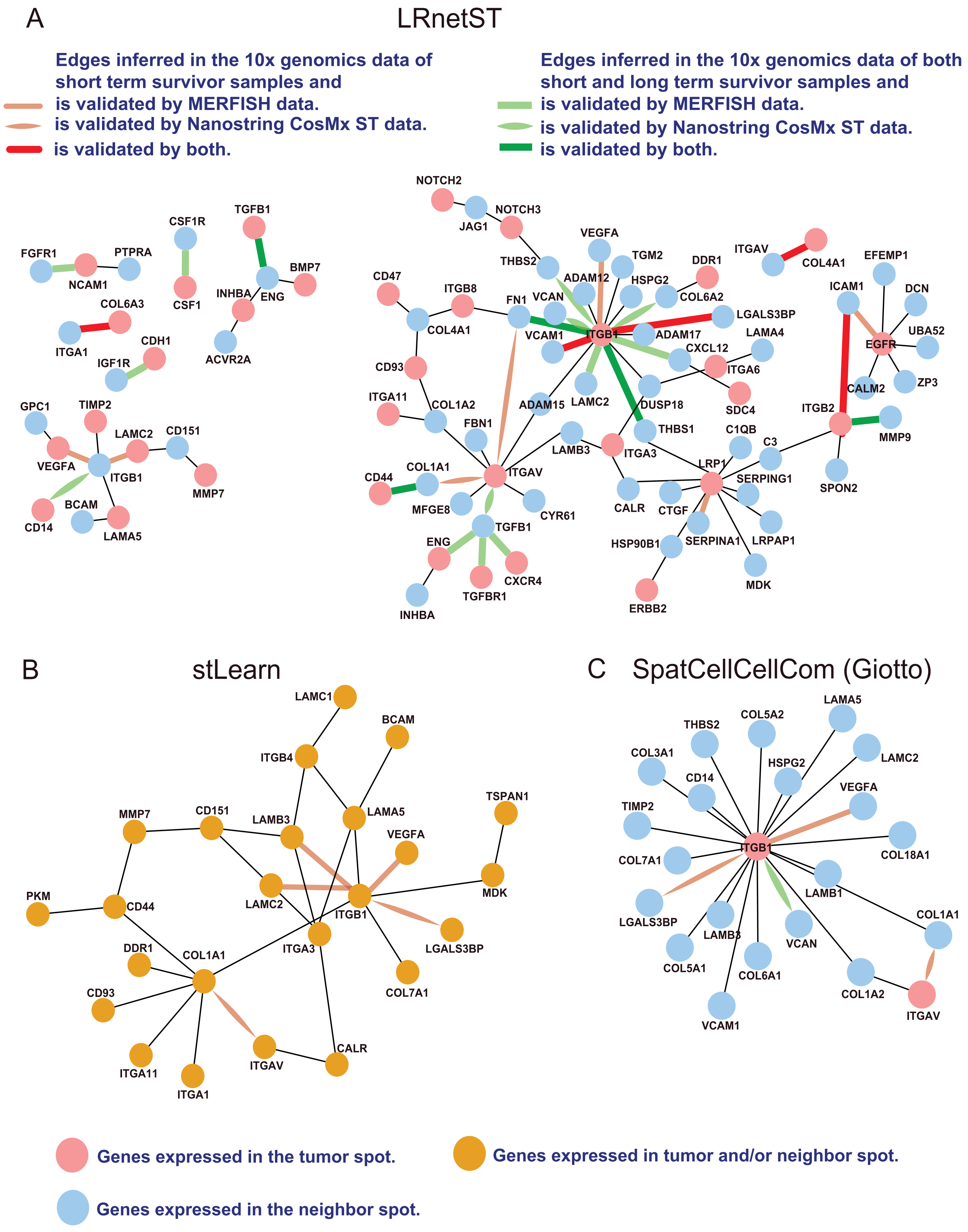


Figure 5. LR network topology showing edges inferred in two or more tumors based on the 10x ST data, with illustration for validation results based on MERFISH and Nanosting CosMx data. (A) Network module topology of LR interactions detected by LRnetST in at least two tumors from the 10x ST data, highlighting modules with at least one validated edge in MERFISH and/or CosMx SMI data; red edges represents interactions inferred only in the short-term survivor patients, while green edges indicate those identified in both short- and long-term survivors in the 10x ST data. (B) Similar to A but for the results of stLearn. (C) Similar to A but for the results of spatCellCellcom (Giotto).

sets, which were consistently validated across multiple tumor samples in the MERFISH data (Fig. 3B and C, 5A). Unfortunately, LRP1 was not measured in the CosMx SMI data. While the role of LRP1 and SERPINA1 interaction in the context of cancer remains largely unexplored, existing research in cardiovascular studies has highlighted its significance in an anti-inflammatory mechanism [42]. By analogy, one may hypothesize that the interaction between LRP1 in tumor cells and SERPINA1 in stroma cells could contribute to an immunosuppressive effect within the TME.

Interaction between LRP1_Tumor and C1QB_Stroma (pattern recognition molecule of innate immunity) [34] were detected only in the two chemo-sensitive patients, suggesting a potential role of LRP1-C1QB crosstalk network in modulating immune response, which may contribute to the chemo-treatment effectiveness in HGSC patients. Unfortunately, C1QB was not measured in the MERFISH data sets, so validation of this interaction using additional resources is warranted as future studies.

Another hub node detected in all four tumors based on 10x ST data sets was ITGB1, which has been inferred to interact with VEGFA and VCAM1 in the two chemo-refractory patients. Interactions between ITGB1 and VEGFA/VCAM1 was validated in either one or both single-cell level data. Intercellular communication mediated via interaction between VEGFA and ITGB1 has been reported to contribute to development, differentiation, and inflammation [38]. This interaction underlies the crosstalk between epithelial cells and fibroblasts relating to tumor inflammation in Pancreatic ductal cancer [43]. While VCAM1 mediates intercellular adhesion by specific binding to the integrin ITGB1 on leukocytes [39], the interaction between ITGB1 and VCAM1 has also been reported in the context of cell migration between brain endothelial cells and central memory T cells [44]. However, our analysis, for the first time, suggests the potential roles of the cross-talk between ITGB1, VEGFA, and/or VCAM1 among ovarian tumors.

This study focuses on analyzing static ST data from single time points. In the future, when temporal ST data are more accessible, causal relationships might be inferred with greater confidence by utilizing dynamic network-building techniques such as temporal link prediction [45, 46].

LR interaction-driven cell–cell communication is a key component of spatial domain communication. Further evaluation of how LR interactions vary across different spatial regions would be valuable and warrants future studies.

Conclusion

ST data offer an unprecedented opportunity to unravel the molecular mechanisms underlying cell–cell interactions within the TME. LRnetST proves to be a valuable tool for constructing LR interaction networks based on ST data. The outcomes of these enhanced analyses promise to unveil crucial contributors driving cell–cell interactions, potentially leading to the identification of new predictive biomarkers or therapeutic targets.

Key Points

- We propose a new method, LRnetST, to characterize the LR interactions among neighboring cells using multicellular or single-cellular ST data.
- LRnetST utilizes directed acyclic graph models with customized treatment to handle the zero-inflated distribution of the ST data and employs an aggregation framework to enhance network inference.
- LRnetST is adaptable to incorporate prior information such as those from existing LR databases.
- Applying LRnetST to both multicellular 10x ST data and independent single-cell MERFISH data and CosMx SMI data of ovarian tumors, we identified several common LR interactions, and demonstrated that LRnetST leads to more reproducible results compared with alternative methods.

Author contributions

Shrabanti Chowdhury (Writingoriginal draft, Formal Analysis, Visualization, Investigation), Sammy Ferri-Borgogno (Formal Analysis, Visualization, Investigation), Peng Yang (Writingoriginal draft, Formal Analysis, Visualization, Investigation), Wenyi Wang (Writingoriginal draft, Formal Analysis, Visualization, Investigation, Supervision, Funding acquisition), Jie Peng (Writingoriginal draft, Formal Analysis, Visualization, Investigation, Supervision, Funding acquisition), Samuel C. Mok (Formal Analysis, Visualization, Investigation, Supervision, Funding acquisition), and Pei Wang (Writingoriginal draft, Formal Analysis, Visualization, Investigation, Supervision, Funding acquisition).

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Conflict of interest: No competing interest is declared.

Funding

This work is supported by grants U24CA271114, U01CA271407, U24CA210993, U01CA214172, U01CA294459, and R01CA268380 from National Institute of Health and National Science Foundation (Division of Mathematical Sciences) grant 1915894. We also thank Dr. Guocheng Yuan and his team members for their instruction on the use of Giotto.

Data availability

An R package of LRnetST is made available as a github repository https://github.com/jie108/LRnetST. All the codes along with the data will be made available through this github repository.

References

1. Cancer Genome Atlas Research Network *et al.* Integrated genomic analyses of ovarian carcinoma. *Nature* 2011;**474**: 609–15.
2. Patch AM, Christie EL, Etemadmoghadam D. *et al.* Whole-genome characterization of chemo-resistant ovarian cancer. *Nature* 2015;**521**:489–94. https://doi.org/10.1038/nature14410
3. Tothill RW, Tinker AV, George J. *et al.* Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res* 2008;**14**:5198–208. https://doi.org/10.1158/1078-0432.CCR-08-0196

4. Zhang H, Liu T, Zhang Z. *et al.* Integrated proteogenomic characterization of human high-grade serous ovarian cancer. *Cell* 2016;**166**:755–65. https://doi.org/10.1016/j.cell.2016.05.069
5. Chowdhury S, Kennedy JJ, Ivey IG. *et al.* Proteogenomic analysis of chemo-refractory high-grade serous ovarian cancer. *Cell* 2023;**186**:3476–3498.e35. https://doi.org/10.1016/j.cell.2023.07.004
6. Zhang B, Chen F, Xu Q. *et al.* Revisiting ovarian cancer microenvironment: A friend or a foe? *Protein Cell* 2018;**9**:674–92. https://doi.org/10.1007/s13238-017-0466-7
7. Tsujikawa T, Mitsuda J, Ogi H. *et al.* Prognostic significance of spatial immune profiles in human solid cancers. *Cancer Sci* 2020;**111**:3426–34. https://doi.org/10.1111/cas.14591
8. Efremova M, Vento-Tormo M, Teichmann SA. *et al.* Cellphonedb: Inferring cell-cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat Protoc* 2020;**15**:1484–506. https://doi.org/10.1038/s41596-020-0292-x
9. Cabello-Aguilar S, Alame M, Kon-Sun-Tack F. *et al.* Single-cellsignalr: Inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res* 2020;**48**:e55. https://doi.org/10.1093/nar/gkaa183
10. Jin S. *et al.* Inference and analysis of cell-cell communication using cellchat. *Nat Commun* 2021;**12**:1–20.
11. Noel F, Massenet-Regad L, Carmi-Levy I. *et al.* Dissection of intercellular communication using the transcriptome-based framework icellnet. *Nat Commun* 2021;**12**:1089. https://doi.org/10.1038/s41467-021-21244-x
12. Nagai JS, Leimkühler NB, Schaub MT. *et al.* Crosstalker: Analysis and visualization of ligand–receptor networks. *Bioinformatics* 2021;**37**:4263–5. https://doi.org/10.1093/bioinformatics/btab370
13. Browaeys R, Saelens W, Saeys W. Nichenet: Modeling intercellular communication by linking ligands to target genes. *Nat Methods* 2020;**17**:159–62. https://doi.org/10.1038/s41592-019-0667-5
14. Cheng J, Zhang J, Wu Z. *et al.* Inferring microenvironmental regulation of gene expression from single-cell rna sequencing data using scmlnet with an application to covid-19. *Brief Bioinform* 2021;**22**:988–1005. https://doi.org/10.1093/bib/bbaa327
15. Cheng J, Zhang J, Wu Z. *et al.* Cytotalk: De novo construction of signal transduction networks using single-cell transcriptomic data. *Sci Adv* 2021;**7**:eabf1356.
16. Asp M, Bergenstråhle J, Lundeberg J. Spatially resolved transcriptomes—Next generation tools for tissue exploration. *Bioessays* 2020;**42**. https://doi.org/10.1002/bies.201900221
17. Marx V. Method of the year: Spatially resolved transcriptomics. *Nat Methods* 2021;**18**:9–14. https://doi.org/10.1038/s41592-020-01033-y
18. Ståhl PL, Salmén F, Vickovic S. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 2016;**353**:78–82. https://doi.org/10.1126/science.aaf2403
19. Dries R, Zhu Q, Dong R. *et al.* Giotto: A toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol* 2021;**22**:1–31.
20. Pham D, Tan X, Balderson B. *et al.* Robust mapping of spatiotemporal trajectories and cell-cell interactions in healthy and diseased tissues. *Nat Commun* 2023;**14**:7739.
21. Tang Z, Zhang T, Yang B. *et al.* Spaci: Deciphering spatial cellular communications through adaptive graph model. *Brief Bioinform* 2023;**24**. https://doi.org/10.1093/bib/bbac563
22. Pearl J. *Causality: Models, Reasoning and Inference*, Vol. **29**. Cambridge Univ Press, 2000.
23. Zhu J, Zhang B, Smith EN. *et al.* Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet* 2008;**40**:854–61. https://doi.org/10.1038/ng.167
24. J. Zhu, P. Sova, Q. Xu, K. M. Dombek, E. Y. Xu, and H., H. Vu, Z. Tu, R. B. Brem, R. E. Bumgarner, E. E. Schadt Vu. Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. *PLoS Biol*, **10**, 2012, e1001301, https://doi.org/10.1371/journal.pbio.1001301.
25. Sung WH, Gong C, Myun-Seok C. *et al.* Estimation of directed acyclic graphs through two-stage adaptive lasso for gene network inference. *J Am Stat Assoc* 2016;**111**:1004–19. https://doi.org/10.1080/01621459.2016.1142880
26. Friedman N, Linial M, Nachman I. *et al.* Using Bayesian networks to analyze expression data. *J Comput Biol* 2000;**7**:601–20. https://doi.org/10.1089/106652700750050961
27. PeÕer D, Regev A, Elidan G. *et al.* Inferring subnetworks from perturbed expression profiles. *Bioinformatics* 2001;**17**:S215–24. https://doi.org/10.1093/bioinformatics/17.suppl_1.S215
28. Sachs K, Perez O, Pe'er D. *et al.* Causal protein-signaling networks derived from multiparameter single-cell data. *Science Signalling* 2005;**308**:523–9. https://doi.org/10.1126/science.1105809
29. Chowdhury S, Wang R, Yu Q. *et al.* Dagbagm: Learning directed acyclic graphs of mixed variables with an application to identify protein biomarkers for treatment response in ovarian cancer. *BMC Bioinformatics* 2022;**23**:321. https://doi.org/10.1186/s12859-022-04864-y
30. Scutari M. Learning Bayesian networks with the bnlearn r package. *J Stat Softw* 2010;**35**. https://doi.org/10.18637/jss.v035.i03
31. Ferri-Borgogno S, Zhu Y, Sheng J. *et al.* Spatial transcriptomics depict ligand–receptor cross-talk heterogeneity at the tumor-stroma interface in long-term ovarian cancer survivors. *Cancer Res* 2023;**83**:1503–16. https://doi.org/10.1158/0008-5472.CAN-22-1821
32. *Vizgen Merfish Ffpe Human Immuno-Oncology Data Set*. Vizgen, 2022.
33. Denisenko E, de Kock L, Tan A. *et al.* Spatial transcriptomics reveals discrete tumour microenvironments and autocrine loops within ovarian cancer subclones. *Nat Commun* 2024;**15**:2860. https://doi.org/10.1038/s41467-024-47271-y
34. Ramilowski J, Goldberg T, Harshbarger J. *et al.* A draft network of ligand–receptor-mediated multicellular signalling in human. *Nat Commun* 2015;**6**:755–65. https://doi.org/10.1038/ncomms8866
35. He X, Zhang J. Why do hubs tend to be essential in protein networks? *PLoS Genet* 2006;**2**. https://doi.org/10.1371/journal.pgen.0020088
36. Zhou W, Ma J, Zhao H. *et al.* Serum exosomes from epithelial ovarian cancer patients contain lrp1, which promotes the migration of epithelial ovarian cancer cell. *Mol Cell Proteomics* 2023;**22**:100520. https://doi.org/10.1016/j.mcpro.2023.100520
37. Strickland DK, Muratoglu SC, Antalis TM. Serpin–enzyme receptors: Ldl receptor-related protein 1. *Methods Enzymol* 2011;**499**:17–31. https://doi.org/10.1016/B978-0-12-386471-0.00002-X
38. R. Miao, X. Dong, J. Gong, Y. Li, X. Guo, and J., J. Wang, Q. Huang, Y. Wang, J. Li, S. Yang, T. Kuang, M. Liu, J. Wan, Z. Zhai, J. Zhong, Y. Yang Wang. Examining the development of chronic thromboembolic pulmonary hypertension at the single-cell level. *Hypertension*, 2022;**79**:562–74. https://doi.org/10.1161/HYPERTENSIONAHA.121.18105.
39. Wu J, Chen ZP, Shang AQ. *et al.* Systemic bioinformatics analysis of recurrent aphthous stomatitis gene expression profiles. *Oncotarget* 2017;**8**:111064–72. https://doi.org/10.18632/oncotarget.22347

40. Lillis AP, Van Duyn LB, Murphy-Ullrich JE. *et al.* Ldl receptor-related protein 1: Unique tissue-specific functions revealed by selective gene knockout studies. *Physiol Rev* 2008;**88**:887–918. https://doi.org/10.1152/physrev.00033.2007
41. Xing P, Liao Z, Ren Z. *et al.* Roles of low-density lipoprotein receptor-related protein 1 in tumors. *Chin J Cancer* 2016;**35**:1–8. https://doi.org/10.1186/s40880-015-0064-0
42. Potere N, Del Buono MG, Mauro AG. *et al.* Low density lipoprotein receptor-related protein-1 in cardiac inflammation and infarct healing. *Frontiers Cardiovasc Med* 2019;**6**. https://doi.org/10.3389/fcvm.2019.00051
43. Kinny-Köster B, Guinn S, Tandurella JA. *et al.* Inflammatory signaling in pancreatic cancer transfers between a single-cell RNA sequencing atlas and co-culture. bioRxiv, 2022. https://doi.org/10.1101/2022.07.14.500096
44. Zhang X, Wang R, Chen H. *et al.* Aged microglia promote peripheral t cell infiltration by reprogramming the microenvironment of neurogenic niches. *Immun Ageing* 2022;**19**:34. https://doi.org/10.1186/s12979-022-00289-6
45. Ma X, Tan S, Xie X. *et al.* Joint multi-label learning and feature extraction for temporal link prediction. *Pattern Recognition* 2022;**121**:108216. https://doi.org/10.1016/j.patcog.2021.108216
46. Ma X, Sun P, Wang Y. Graph regularized nonnegative matrix factorization for temporal link prediction in dynamic networks. *Physica A* 2018;**496**:121–36. https://doi.org/10.1016/j.physa.2017.12.092

© The Author(s) 2025. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Briefings in Bioinformatics, 2025, **26(2)**, bbaf085
https://doi.org/10.1093/bib/bbaf085
Case Study